PREDIKSI RISIKO BURNOUT KARYAWAN MENGGUNAKAN K-MEANS

Muhamad Vicky Oktafrian, Raihan Ramadhan, Anggraini Puspita Sari*
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Surabaya, Jawa Timur, (031) 8737450
22081010028@student.upnjatim.ac.id

ABSTRAK

Burnout merupakan masalah serius dalam dunia kerja yang berdampak pada kesehatan mental karyawan serta kinerja organisasi. Tingginya beban kerja, lembur, dan rendahnya kualitas hubungan kerja menjadi faktor yang berkontribusi terhadap meningkatnya risiko burnout. Penelitian ini bertujuan untuk mengelompokkan tingkat risiko burnout karyawan menggunakan algoritma K-Means Clustering berdasarkan empat variabel utama, yaitu usia, jenis kelamin, riwayat lembur, dan kepuasan hubungan dengan rekan kerja. Metode penelitian meliputi pengumpulan data karyawan sebanyak 1.480 data, pra-pemrosesan data berupa encoding variabel kategorik dan normalisasi menggunakan Min-Max Scaler, serta pemodelan clustering menggunakan algoritma K-Means. Penentuan jumlah klaster optimal dilakukan dengan Elbow Method dan Silhouette Score. Hasil evaluasi menunjukkan bahwa jumlah klaster optimal adalah empat dengan nilai Silhouette Score sebesar 0,5472. Keempat klaster tersebut merepresentasikan tingkat risiko burnout yang berbeda, yaitu Non-Burnout, Low-Burnout, Medium-Burnout, dan High-Burnout. Hasil analisis menunjukkan bahwa karyawan yang sering lembur dan memiliki tingkat kepuasan hubungan kerja yang rendah cenderung berada pada kelompok risiko burnout yang lebih tinggi. Penelitian ini diharapkan dapat membantu organisasi dalam mendeteksi risiko burnout secara dini sebagai dasar pengambilan keputusan dan perancangan strategi intervensi yang lebih tepat sasaran.

Kata kunci : *Burnout, Karyawan, K-Means, Clustering, Risiko Kerja.*

1. PENDAHULUAN

Burnout merupakan salah satu permasalahan serius yang dihadapi oleh banyak individu sebagai karyawan atau pekerja dalam organisasi di berbagai sektor industri. Kondisi ini ditandai oleh kelelahan emosional yang berkepanjangan, munculnya rasa sinisme dan ketidaknyamanan terhadap pekerjaan, serta menurunnya pencapaian pribadi [1]. Burnout sering kali muncul akibat tekanan kerja yang berlebihan dan kurangnya dukungan sosial atau penghargaan di tempat kerja [2], [3]. Dampaknya tidak hanya dirasakan oleh individu, tetapi juga memengaruhi kinerja dan stabilitas tenaga kerja dalam organisasi secara keseluruhan [4].

Faktor-faktor yang dapat memengaruhi burnout sangat beragam, mulai dari lingkungan kerja yang tidak mendukung, beban kerja yang tinggi, hingga karakteristik demografis seperti usia, jenis kelamin, dan pengalaman kerja [5]. Lingkungan kerja yang buruk dan beban kerja berlebih memiliki hubungan signifikan dengan tingkat burnout pada karyawan. Karyawan yang bekerja dalam sistem organisasi yang kurang adaptif terhadap kesejahteraan mental cenderung mengalami burnout lebih cepat [6]. Oleh karena itu, intervensi perlu dilakukan dengan tidak hanya berfokus pada kinerja, tetapi juga pada manajemen psikososial [7].

Organisasi yang hanya berfokus pada produktivitas tanpa memperhatikan kesehatan mental dan emosional karyawan akan menghadapi risiko jangka panjang, seperti penurunan motivasi, absensi tinggi, hingga turnover. Kondisi ini menunjukkan bahwa upaya pencegahan burnout perlu menjadi bagian integral dari strategi manajemen sumber daya manusia. Dengan semakin besarnya data dalam sistem administrasi SDM, kebutuhan untuk menerapkan pendekatan analitik berbasis data menjadi lebih signifikan. Teknologi informasi memungkinkan organisasi untuk memproses data karyawan dalam jumlah besar secara efisien [8]. Metode analisis prediktif seperti machine learning telah digunakan secara luas dalam berbagai konteks diagnosis dan deteksi dini, termasuk dalam isu kesehatan mental [9].

Salah satu algoritma populer dalam unsupervised learning adalah K-Means, yang bekerja berdasarkan prinsip pengelompokan data terhadap centroid terdekat untuk meminimalkan jarak antar titik dalam cluster [10]. Kelebihan K-Means meliputi kemudahan implementasi, efisiensi komputasi, serta kemampuan menangani data numerik dalam jumlah besar dengan cepat. Algoritma ini juga bersifat skalabel dan mudah diintegrasikan ke sistem informasi perusahaan [11], [12]. Namun, K-Means juga memiliki keterbatasan, seperti ketergantungan pada inisialisasi centroid awal, kesulitan dalam menentukan jumlah cluster (K) secara optimal, dan sensitivitas terhadap outlier atau bentuk cluster yang tidak bulat [13], [14], [15]. Untuk mengatasi kekurangan tersebut, sejumlah optimasi seperti teknik inisialisasi K-Means++, random projection, atau penggunaan divergence measures seperti Bregman divergence dapat diterapkan guna meningkatkan akurasi dan konsistensi hasil clustering.

Berbagai studi terdahulu telah menerapkan algoritma K-Means dalam beragam domain non-klinis, seperti pengelompokan pasar ekspor, analisis perilaku pengguna e-commerce, hingga prediksi risiko dan komplikasi penyakit berbasis data mining [16], [17], [18]. Namun demikian, penerapan metode

clustering untuk mengidentifikasi dan mengelompokkan tingkat risiko burnout karyawan berdasarkan faktor demografis serta karakteristik kerja masih relatif terbatas. Hal ini menunjukkan adanya celah penelitian yang perlu dikaji lebih lanjut, khususnya dalam konteks pemanfaatan data organisasi untuk mendukung pengambilan keputusan terkait kesejahteraan karyawan.

Pendekatan unsupervised learning seperti K-Means memungkinkan pemrosesan data tanpa memerlukan label diagnosis awal, sehingga sesuai untuk menemukan pola tersembunyi dalam data burnout [19]. Variabel-variabel yang digunakan dalam penelitian ini telah terbukti berhubungan signifikan dengan kelelahan kerja dan kesehatan mental dalam berbagai studi sebelumnya [20]. Keterbaruan penelitian ini terletak pada penerapan algoritma K-Means untuk identifikasi dini risiko burnout berbasis data organisasi, yang masih jarang dikaji dalam penelitian sejenis. Melalui tahapan prapemrosesan data, proses clustering, serta evaluasi hasil menggunakan metode Elbow Curve dan Silhouette Score, penelitian ini bertujuan untuk menghasilkan pengelompokan risiko burnout yang objektif. Hasil penelitian diharapkan dapat membantu manajemen sumber daya manusia dalam mendeteksi risiko burnout secara lebih dini dan merancang strategi intervensi yang tepat sasaran berbasis data.

2. TINJAUAN PUSTAKA

Perkembangan penelitian mengenai burnout dalam dua dekade terakhir menunjukkan bahwa kelelahan kerja telah menjadi isu utama dalam manajemen sumber daya manusia modern. Burnout tidak lagi dipandang sebagai masalah individu semata, melainkan sebagai fenomena organisasi yang berdampak langsung terhadap produktivitas dan stabilitas kerja. Maslach dan Leiter menyatakan bahwa burnout ditandai oleh kelelahan emosional, sikap sinis terhadap pekerjaan, serta menurunnya pencapaian personal [1]. Temuan ini diperkuat oleh Schaufeli et al. yang menegaskan bahwa burnout muncul sebagai akibat dari ketidakseimbangan antara tuntutan kerja dan kapasitas psikologis karyawan [2].

Dalam konteks faktor penyebab, Bakker dan Demerouti mengemukakan bahwa tingginya beban kerja, tekanan waktu, serta rendahnya dukungan sosial merupakan determinan utama munculnya burnout [3]. Halbesleben dan Buckley juga menunjukkan bahwa karakteristik demografis seperti usia dan jenis kelamin berpengaruh terhadap tingkat kerentanan burnout [4]. Karyawan usia muda cenderung lebih rentan terhadap stres kerja akibat keterbatasan pengalaman dalam mengelola tekanan profesional.

Selain faktor individu, lingkungan kerja memiliki peran signifikan dalam membentuk tingkat kelelahan karyawan. Leiter dan Maslach menemukan bahwa sistem organisasi yang kurang mendukung kesejahteraan psikologis mempercepat munculnya burnout [5]. Sonnentag menambahkan bahwa hubungan kerja yang buruk dan minimnya penghargaan memperkuat dampak negatif tekanan kerja terhadap kesehatan mental [6]. Penelitian lain juga menunjukkan bahwa rendahnya kepuasan hubungan dengan rekan kerja berkorelasi dengan meningkatnya tingkat stres dan kelelahan emosional [7].

Seiring meningkatnya kompleksitas sistem organisasi, pemanfaatan teknologi informasi dalam pengelolaan sumber daya manusia semakin berkembang. Chen et al. menyatakan bahwa analisis data berbasis teknologi memungkinkan organisasi untuk memetakan kondisi psikologis karyawan secara lebih sistematis [8]. Data mining menjadi salah satu pendekatan yang banyak digunakan untuk mengekstraksi pola tersembunyi dari data karyawan dalam jumlah besar [9]. Pendekatan ini mendukung pengambilan keputusan yang lebih objektif dan berbasis bukti.

Dalam bidang machine learning, K-Means Clustering merupakan salah satu algoritma unsupervised learning yang paling banyak digunakan untuk pengelompokan data. Menurut Jain, algoritma ini bekerja dengan meminimalkan jarak antar data dalam satu klaster melalui penentuan centroid optimal [10]. K-Means dikenal memiliki keunggulan dalam efisiensi komputasi dan kemudahan implementasi, sehingga banyak diterapkan dalam sistem informasi organisasi [11].

Namun, sejumlah penelitian juga menyoroti keterbatasan algoritma K-Means. Aggarwal menjelaskan bahwa hasil clustering sangat bergantung pada inisialisasi awal centroid dan penentuan jumlah cluster [12]. Selain itu, K-Means cenderung sensitif terhadap outlier dan kurang efektif pada data dengan distribusi tidak homogen [13]. Oleh karena itu, diperlukan metode evaluasi tambahan untuk memastikan kualitas hasil clustering.

Metode Elbow dan Silhouette Score merupakan dua teknik yang umum digunakan untuk menentukan jumlah cluster optimal. Rousseeuw menyatakan bahwa Silhouette Score mampu mengukur tingkat kohesi dan separasi antar cluster secara kuantitatif [14]. Sementara itu, Elbow Method digunakan untuk mengidentifikasi titik keseimbangan antara kompleksitas model dan kualitas pengelompokan [15]. Kombinasi kedua metode ini dapat meningkatkan reliabilitas hasil clustering.

Dalam konteks burnout, penerapan metode clustering masih relatif terbatas. Beberapa penelitian sebelumnya menggunakan K-Means untuk mengelompokkan tingkat stres kerja dan kepuasan karyawan [16]. Penelitian oleh Kumar et al. menunjukkan bahwa clustering mampu mengidentifikasi kelompok karyawan berisiko tinggi burnout secara lebih dini [17]. Namun, sebagian besar studi masih berfokus pada variabel terbatas dan belum mengintegrasikan faktor demografis serta sosial secara komprehensif.

Secara keseluruhan, tinjauan pustaka menunjukkan bahwa burnout dipengaruhi oleh interaksi kompleks antara faktor individu, lingkungan kerja, dan sistem organisasi. Di satu sisi, pemanfaatan data mining dan machine learning membuka peluang untuk deteksi dini risiko burnout secara objektif. Di sisi lain, keterbatasan algoritma clustering menuntut pemilihan metode dan evaluasi yang cermat. Oleh karena itu, penelitian ini menempatkan K-Means Clustering sebagai pendekatan analitis untuk menyintesiskan pola burnout karyawan berdasarkan usia, jenis kelamin, riwayat lembur, dan kepuasan hubungan kerja guna menghasilkan pemahaman yang lebih komprehensif.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode clustering untuk mengelompokkan risiko burnout pada karyawan. Metodologi yang digunakan dalam penelitian dijabarkan pada Gambar 1, ini meliputi beberapa tahapan, yaitu: pengumpulan dataset, pra-pemrosesan dataset, penerapan clustering algoritma K-Means, evaluasi model clustering, serta analisis hasil.

Gambar 1. Diagram Alir Penelitian

3.1. Pengumpulan Data

Dataset yang digunakan bersumber dari Kaggle dengan total 1.480 data karyawan. Variabel yang digunakan dalam penelitian ini terdiri dari empat faktor utama yang berkaitan dengan burnout: usia (Age), jenis kelamin (Gender), riwayat lembur (OverTime), dan kepuasan hubungan dengan rekan kerja (RelationshipSatisfaction). Pemilihan variabel ini didasarkan pada kajian literatur yang menunjukkan korelasi signifikan dengan tingkat burnout.

3.2. Pra-Pemrosesan

Pada tahapan ini, pra-pemrosesan data terhadap fitur yang akan digunakan diantaranya variabel Age, Gender, Overtime, dan RelationshipSatisfaction meliputi pengecekan missing values, encoding variabel kategorik, dan normalisasi data. Pengecekan nilai null dilakukan untuk menghindari ketidak seimbangan ketika model dibangun, kemudian dilakukannya encoding menjadi bentuk numerik pada variabel Gender dan Overtime menggunakan label encoding agar dapat digunakan dalam algoritma K-Means. Terakhir adalah melakukan normalisasi data menggunakan metode MinMaxScaller dengan tujuan menghindari bias terhadap fitur yang memiliki skala besar dan menstandarkan nilai fitur dalam rentang 0 hingga 1[16]. Metode Min-Max Scaling menggunakan persamaan berikut:

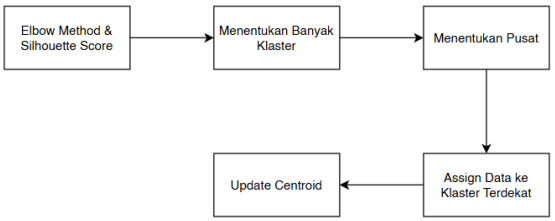
$$X_{Scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

dimana,
X : Nilai asli dari fitur yang ingin dinormalisasi.
Xmin : Nilai minimum dari fitur tersebut.
Xmax : Nilai maksimum dari fitur tersebut.
Xscaled : Nilai hasil normalisasi, yaitu hasil dari skala baru dalam rentang 0 hingga 1.

3.3. Clustering Menggunakan K-Means

Proses pemodelan dan evaluasi dilakukan menggunakan algoritma K-Means Clustering. Tahapan ini bertujuan untuk mengelompokkan data karyawan berdasarkan kemiripan karakteristik tertentu yang berkaitan dengan potensi burnout. Sebelum algoritma K-Means dijalankan, jumlah klaster optimal terlebih dahulu ditentukan menggunakan dua metode evaluasi, yaitu Elbow Method dan Silhouette Score. Elbow Method digunakan untuk mengevaluasi nilai inertia (jumlah kuadrat jarak antar data dengan pusat klasternya), sedangkan Silhouette Score mengukur seberapa baik setiap objek cocok dengan klasternya sendiri dibandingkan dengan klaster lain.

Setelah jumlah klaster (nilai k) ditentukan, proses K-Means dimulai dengan inisialisasi centroid secara acak. Selanjutnya, setiap data akan dihitung jaraknya ke masing-masing centroid dan akan dikelompokkan ke klaster dengan jarak terdekat. Proses ini diikuti dengan pembaruan posisi centroid berdasarkan rata-rata dari anggota klaster baru. Langkah tersebut terus diulang hingga posisi centroid tidak lagi berubah (konvergen). Alur proses ini secara ringkas dapat dilihat pada Gambar 2 berikut:



Gambar 2. Diagram Alir Model K-Means

3.4. Analisis Hasil

Hasil klasterisasi yang diperoleh dari algoritma K-Means kemudian dianalisis untuk mengidentifikasi kelompok karyawan berdasarkan tingkat burnout. Setiap klaster ditafsirkan berdasarkan karakteristik empat faktor utama. Analisis ini bertujuan untuk memahami profil masing-masing kelompok karyawan dan mendeteksi pola-pola yang mengindikasikan tingkat burnout yang berbeda.

4. HASIL DAN PEMBAHASAN

4.1. Pra-Pemrosesan

Setelah dilakukan pengecekan pada missing values untuk empat faktor utama, tidak ditemukan adanya nilai kosong pada data. Hal ini ditunjukkan pada Gambar 3 yang menampilkan hasil dari scanning.

```
=== STEP 2A: BASIC DATA INFORMATION ===
Missing values:
Age                       0
Gender                    0
RelationshipSatisfaction  0
OverTime                  0
dtype: int64
```

Gambar 3. Hasil Pengecekan Nilai Null Pada Empat Variabel

Karena tidak terdapat missing values maka dilakukan encoding pada data kategorikal, dalam kasus ini yaitu Gender dan Overtime. Hasil encoding ditampilkan pada Gambar 4, langkahnya adalah mengubah variabel Male menjadi nilai 1 dan Female menjadi nilai 0 serta mengubah variabel Overtime Yes menjadi 1 dan No menjadi 0.

```
# Encode Gender (Male/Female menjadi 0/1)
burnout_processed['Gender'] = le_gender.fit_transform(burnout_processed['Gender'])
# Encode OverTime (No/Yes menjadi 0/1)
burnout_processed['OverTime'] = le_overtime.fit_transform(burnout_processed['OverTime'])
```

```
Data setelah encoding:
   Age  Gender  RelationshipSatisfaction  OverTime
0   18       1                         3         0
1   18       0                         1         0
2   18       1                         4         1
3   18       1                         4         0
4   18       1                         4         0
```

Gambar 4. Proses Encoding Pada Data Kategorikal

Langkah terakhir dalam tahap pra-pemrosesan data adalah melakukan normalisasi untuk menghindari dominasi fitur-fitur yang memiliki skala nilai yang besar. Misalnya, pada fitur Age yang memiliki rentang nilai antara 18 hingga 55, nilai terendah yaitu 18 akan dikonversi menjadi 0 setelah dinormalisasi menggunakan persamaan (1). Sementara itu, pada fitur RelationshipSatisfaction yang memiliki nilai antara 1 hingga 4, suatu nilai sebesar 3 akan dinormalisasi menjadi sekitar 0,67 berdasarkan persamaan (1). Gambar 5 menunjukkan hasil akhir dari proses normalisasi menggunakan MinMaxScaler pada dataset yang telah melalui tahapan encoding sebelumnya.

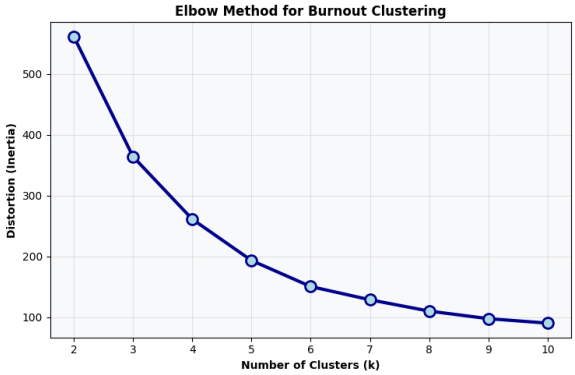
```
Data setelah normalisasi dengan MinMaxScaler:
   Age  Gender  RelationshipSatisfaction  OverTime
0  0.0     1.0                  0.666667       0.0
1  0.0     0.0                  0.000000       0.0
2  0.0     1.0                  1.000000       1.0
3  0.0     1.0                  1.000000       0.0
4  0.0     1.0                  1.000000       0.0
```

Gambar 5. Hasil Normalisasi Dataset

4.2. Clustering Menggunakan K-Means

Setelah dilakukan pra-pemrosesan dataset, langkah berikutnya adalah menentukan jumlah cluster optimal untuk pengelompokan data karyawan berdasarkan risiko burnout. Untuk menentukan jumlah cluster yang tepat, digunakan dua metode analisis yaitu Metode Elbow dan Silhouette Analysis [20].

Metode Elbow dilakukan dengan menghitung nilai inertia untuk berbagai jumlah cluster, dalam hal ini 2 sampai 10. Hasil dari perhitungan ini divisualisasikan dalam sebuah grafik yang menunjukkan perubahan nilai inertia terhadap jumlah cluster. Titik dimana penurunan mulai melambat dianggap sebagai cluster yang optimal. Inilah yang dinamakan Elbow Curve yang ditampilkan pada Gambar 6. Berdasarkan metode Elbow yang ditunjukkan pada Gambar tersebut, jumlah cluster yang optimal untuk pengelompokan burnout karyawan adalah k = 4. Nilai ini memberikan keseimbangan antara akurasi segmentasi dan kompleksitas model.



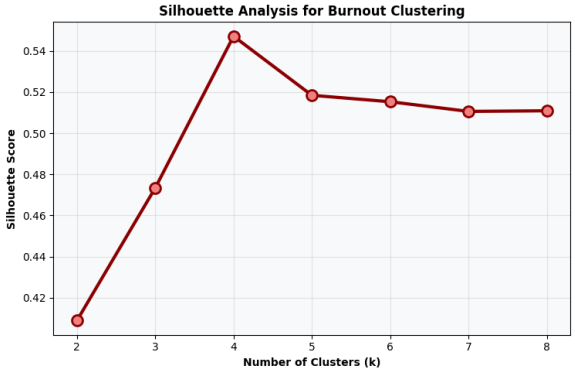
Gambar 6. Hasil Elbow Method

Sementara itu, Silhouette Analysis digunakan untuk mengevaluasi seberapa baik data dikelompokkan ke dalam cluster. Nilai silhouette score dihitung untuk jumlah cluster dari 2 hingga 8, di mana skor yang lebih tinggi menunjukkan pemisahan cluster yang lebih baik dan lebih koheren. Hasilnya divisualisasikan dalam Gambar 7.

```
Silhouette Scores:
k=2: 0.4088
k=3: 0.4734
k=4: 0.5472
k=5: 0.5184
k=6: 0.5153
k=7: 0.5106
k=8: 0.5109
k=9: 0.4755
k=10: 0.4837
```

Gambar 7. Analisis Silhouette Score

Dari hasil tersebut, terlihat bahwa nilai silhouette tertinggi diperoleh pada k = 4 dengan skor 0.5472, menunjukkan kualitas klasterisasi terbaik dibanding jumlah cluster lainnya. Kemudian divisualisasikan dalam bentuk grafik yang ditampilkan pada Gambar 8.



Gambar 8. Grafik Silhouette Score

Dari kedua analisis ini, dipilih jumlah cluster optimal sebanyak empat (k=4), karena memberikan keseimbangan antara interpretabilitas dan performa clustering.

4.3. Analisis Hasil

Setelah jumlah cluster ditentukan, algoritma K-Means Clustering diterapkan untuk membagi data ke dalam empat cluster. Setiap cluster kemudian dianalisis lebih lanjut berdasarkan karakteristik fitur penting seperti usia (Age), lembur (OverTime), tingkat kepuasan hubungan (RelationshipSatisfaction), dan jenis kelamin (Gender). Untuk menilai tingkat kelelahan kerja (burnout), digunakan formula sederhana: persentase lembur ditambahkan dengan penalti dari rendahnya tingkat kepuasan hubungan (dalam skala 1–4). Misalnya, jika kepuasan adalah 1 (terendah), maka skor burnout akan lebih tinggi.

Gambar 9 menunjukkan distribusi karakteristik karyawan berdasarkan hasil clustering tingkat risiko burnout. Total jumlah data sebanyak 1480 dengan rincian cluster 0 berjumlah 180, cluster 1 berjumlah 651, cluster 2 berjumlah 411, dan cluster 3 berjumlah 238 menunjukkan bahwa Cluster 0 dan 1 memiliki risiko burnout rendah karena seluruh anggotanya tidak melakukan lembur, sedangkan Cluster 2 dan 3 memiliki risiko burnout menengah karena seluruh anggotanya melakukan lembur meskipun memiliki kepuasan hubungan kerja yang relatif baik. Jumlah total data ini sinkron dengan data yang diproses sebelumnya.

```
Cluster 0:
  Jumlah: 180
  Persentase lembur: 100.00%
  Rata-rata RelationshipSatisfaction: 2.70
  Risk Level: Medium

Cluster 1:
  Jumlah: 651
  Persentase lembur: 0.00%
  Rata-rata RelationshipSatisfaction: 2.68
  Risk Level: Low

Cluster 2:
  Jumlah: 411
  Persentase lembur: 0.00%
  Rata-rata RelationshipSatisfaction: 2.67
  Risk Level: Low

Cluster 3:
  Jumlah: 238
  Persentase lembur: 100.00%
  Rata-rata RelationshipSatisfaction: 2.86
  Risk Level: Medium
```

Gambar 9. Distribusi Karakteristik

Hasil analisis setiap cluster dijabarkan dalam sebuah tabel (Tabel 1 – Hasil Clustering), yang menunjukkan jenis kelamin, rata-rata usia, persentase lembur, tingkat kepuasan, dan skor burnout untuk masing-masing klaster. Berdasarkan skor ini, setiap klaster dikategorikan ke dalam level burnout: Non Burnout, Low Burnout, Medium Burnout, dan High Burnout. Kategori ini ditentukan dengan mengurutkan cluster dari yang memiliki skor burnout terendah hingga tertinggi.

Tabel 1. Hasil Clustering

Cluster	Jumlah	Karakteristik	Level Burnout
Cluster 0	180	Didominasi oleh Perempuan, rata-rata usia mereka adalah 37 tahun. Sering lembur dan tingkat kepuasan hubungan dengan rekan kerja berada di angka 2.70.	High Burnout
Cluster 1	651	Didominasi oleh Laki-laki, rata-rata usia mereka adalah 36 tahun. Jarang lembur dan tingkat kepuasan hubungan dengan rekan kerja berada di angka 2.68.	Non Burnout
Cluster 2	411	Didominasi oleh Perempuan, rata-rata usia mereka adalah 37,4 tahun. Jarang lembur dan tingkat kepuasan hubungan dengan rekan kerja berada di angka 2.67.	Low Burnout
Cluster 3	238	Didominasi oleh Laki-laki, rata-rata usia mereka adalah 37,5 tahun. Sering lembur dan tingkat kepuasan hubungan dengan rekan kerja berada di angka 2.86.	Medium Burnout

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan model prediksi risiko burnout karyawan menggunakan algoritma K-Means Clustering berdasarkan empat variabel utama, yaitu usia, jenis kelamin, riwayat lembur, dan kepuasan hubungan kerja. Proses pra-pemrosesan data dilakukan secara menyeluruh, mencakup pengecekan nilai kosong, encoding variabel kategorik, dan normalisasi data. Penentuan jumlah cluster optimal menggunakan Elbow Method dan Silhouette Score menghasilkan empat cluster dengan nilai silhouette tertinggi sebesar 0,5472 pada k = 4, yang menunjukkan kualitas klasterisasi yang baik. Hasil analisis menunjukkan bahwa karyawan laki-laki, berusia muda, sering lembur, dan memiliki kepuasan hubungan kerja rendah cenderung berada pada kelompok risiko burnout tinggi. Model ini dapat digunakan organisasi untuk mendeteksi kelompok karyawan rentan burnout secara dini, sehingga intervensi preventif dan strategi peningkatan kesejahteraan serta produktivitas dapat dilakukan secara lebih efektif dan tepat sasaran. Penelitian ini

menegaskan pentingnya pemanfaatan data mining dalam manajemen sumber daya manusia untuk mencegah burnout secara sistematis dan berbasis data.

DAFTAR PUSTAKA

[1] C. Maslach, W. B. Schaufeli, and M. P. Leiter, "Job Burnout," *Annu. Rev. Psychol.*, vol. 52, no. 1, pp. 397–422, Feb. 2022, doi: 10.1146/annurev.psych.52.1.397.

[2] W. B. Schaufeli and A. B. Bakker, "Job demands, job resources, and their relationship with burnout and engagement: a multi-sample study," *J. Organ. Behav.*, vol. 25, no. 3, pp. 293–315, May 2020, doi: 10.1002/job.248.

[3] J. R. B. Halbesleben, "Sources of social support and burnout: A meta-analytic test of the conservation of resources model.," *Journal of Applied Psychology*, vol. 91, no. 5, pp. 1134–1145, Sep. 2021, doi: 10.1037/0021-9010.91.5.1134.

[4] M. P. LEITER and C. MASLACH, "Nurse turnover: the mediating role of burnout," *J. Nurs. Manag.*, vol. 17, no. 3, pp. 331–339, Apr. 2020, doi: 10.1111/j.1365-2834.2009.01004.x.

[5] J. C. Quick and L. E. Tetrick, *Handbook of occupational health psychology*. 2020.

[6] R. Alfikri, Rd. Halim, M. Syukri, L. Nurdini, and F. Islam, "Status Gizi dengan Kelelahan Kerja Karyawan Bagian Proses dan Teknik Pabrik Kelapa Sawit," *Jurnal Kesehatan Komunitas*, vol. 7, no. 3, pp. 271–276, Dec. 2021, doi: 10.25311/keskom.Vol7.Iss3.983.

[7] F. D. Ariyanti and F. Gunawan, "Analysis of Request for Quotation (RFQ) with Rejected Status Use K-Modes and Ward's Clustering Methods. A Case Study of B2B E-Commerce Indotrading.Com," 2023, pp. 643–654. doi: 10.1007/978-3-031-29078-7_56.

[8] R. A. Gulhane and S. R. Gupta, "An Early Disease Prediction and Risk Analysis of Diabetic Mellitus using Electronic Medical Records," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1085, no. 1, p. 012023, Feb. 2021, doi: 10.1088/1757-899X/1085/1/012023.

[9] Sari A P, Prasetya D A, Aditiawan F P, and Haromainy M M A, "Prediksi Gangguan Kesehatan Mental pada Kalangan Mahasiswa Menggunakan Metode Pseudo-Labeling dan Algoritma Regresi Logistik," *TAMIKA: Jurnal Tugas Akhir Manajemen Informatika & Komputerisasi Akuntansi*, vol. 4, no. 2, Dec. 2024.

[10] Y. Liu, "Analysis and Prediction of College Students' Mental Health Based on K-means Clustering Algorithm," *Applied Mathematics and Nonlinear Sciences*, vol. 7, no. 1, pp. 501–512, Jan. 2022, doi: 10.2478/amns.2021.1.00099.

[11] N. A. Baghdadi, S. K. Alsayed, G. A. Malki, H. M. Balaha, and S. M. Farghaly Abdelaliem, "An Analysis of Burnout among Female Nurse Educators in Saudi Arabia Using K-Means Clustering," *Eur. J. Investig. Health Psychol. Educ.*, vol. 13, no. 1, pp. 33–53, Dec. 2022, doi: 10.3390/ejihpe13010003.

[12] Yuliastati Putri Sugiarta Karlim, Aditiya Hermawan, and Ardiane Rossi Kurniawan Maranto, "Clustering Mental Health on Instagram Users Using K-Means Algorithm," *bit-Tech*, vol. 6, no. 1, pp. 32–39, Aug. 2023, doi: 10.32877/bt.v6i1.880.

[13] M. A. P. Subali, I. G. R. A. Sugiartha, K. Budiarta, and I. M. B. Adnyana, "Clustering Balinese Language Documents using the Balinese Stemmer Method and Mini Batch K-Means with K-Means++," *Journal of Applied Informatics and Computing*, vol. 7, no. 2, pp. 258–262, Dec. 2023, doi: 10.30871/jaic.v7i2.6476.

[14] M. E. Celebi, H. A. Kingravi, and P. A. Vela, "A comparative study of efficient initialization methods for the k-means clustering algorithm," *Expert Syst. Appl.*, vol. 40, no. 1, pp. 200–210, Jan. 2023, doi: 10.1016/j.eswa.2012.07.021.

[15] A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh, "Clustering with Bregman Divergences," in *Proceedings of the 2024 SIAM International Conference on Data Mining*, Philadelphia, PA: Society for Industrial and Applied Mathematics, Apr. 2024, pp. 234–245. doi: 10.1137/1.9781611972740.22.

[16] A. M. Hudri Ritonga and Safrizal, "Analisis Clustering Gaji Karyawan Menggunakan K-Means dan Elbow Method Rumah Sakit XYZ," *Jurnal Algoritma*, vol. 22, no. 2, Nov. 2025, doi: 10.33364/algoritma/v.22-2.2915.

[17] U. H. Wanaskar, S. R. Vij, and D. Mukhopadhyay, "A Hybrid Web Recommendation System Based on the Improved Association Rule Mining Algorithm," *Journal of Software Engineering and Applications*, vol. 06, no. 08, pp. 396–404, 2023, doi: 10.4236/jsea.2013.68049.

[18] C.-F. Chien and L.-F. Chen, "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry," *Expert Syst. Appl.*, vol. 34, no. 1, pp. 280–290, Jan. 2021, doi: 10.1016/j.eswa.2006.09.003.

[19] Brown K A, Sarkar I N, Crowley K M, Aluthge D P, and Chen E S, "An Unsupervised Cluster Analysis of Post-COVID-19 Mental Health Outcomes and Associated Comorbidities," *PubMed*, 2022.

[20] C. El Morr *et al.*, "Predictive Machine Learning Models for Assessing Lebanese University Students' Depression, Anxiety, and Stress During COVID-19," *J. Prim. Care Community Health*, vol. 15, Jan. 2024, doi: 10.1177/21501319241235588